Effective sample size approximations as entropy measures

L. Martino[⋆], V. Elvira[◇],

[⋆] Universidad Rey Juan Carlos, Madrid, Spain.

[⊤] University of Edinburgh, UK.

Abstract

In this work, we analyze alternative effective sample size (ESS) measures for importance sampling algorithms. More specifically, we study a family of ESS approximations introduced in [11]. We show that all the ESS functions included in this family (called Huggins-Roy's family) satisfy all the required theoretical conditions introduced in [17]. We also highlight the relationship of this family with the Rényi entropy. By numerical simulations, we study the performance of different ESS approximations introducing also an optimal linear combination of the most promising ESS indices introduced in literature. Moreover, we obtain the best ESS approximation within the Huggins-Roy's family, that provides almost a perfect match with the theoretical ESS values.

Keywords: Importance Sampling; Effective Sample Size; Resampling; Sequential Monte Carlo; Particle filtering.

The effective sample size (ESS) is an important concept in order to measure the efficiency of different Monte Carlo methods, such as Markov Chain Monte Carlo (MCMC) [9, 15] and Importance Sampling (IS) techniques [1, 3]. In an IS context, heuristically speaking, we can assert that ESS measures how many independent identically distributed (i.i.d.) samples, drawn directly from the target distribution $\bar{\pi}(\mathbf{x}) = \frac{1}{Z}\pi(\mathbf{x})$, are equivalent *in some sense* to the N weighted samples, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, drawn from a proposal distribution $q(\mathbf{x})$ and weighted according to the ratio $w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$ [18]. This consideration is represented in the first box of Figure 1, referred as "abstract ESS concept".

The foundational mathematical definition [9, 13] considers the ESS function proportional to the ratio between the variance of the ideal Monte Carlo estimator (drawing samples directly from the target) over the variance of the estimator obtained by the IS technique, using with the same number of samples in both estimators. This definition presents some drawbacks (see [17, 8] for an exhaustive discussion) and is useless for practical purposes since it cannot be computed in general. Hence, approximations of this theoretical formula are required. In Figure 1, this theoretical definition is represented by the second box.

Within a IS context, the most common choice in literature to approximate this theoretical ESS definition is $\widehat{ESS} = \frac{1}{\sum_{n=1}^{M} \bar{w}_n^2}$, which involves (only) the normalized importance weights $\bar{w}_n = \frac{w_n}{\sum_{j=1}^{N} w_j}$, $n = 1, \ldots, N$ [5, 6, 14, 18]. This expression has been widely used in particle filtering in order to apply the resampling steps adaptively [6, 5, 10]. However, it presents different weaknesses since it has been obtained after several approximations of the theoretical definition (see [17] for further details).

Several other approximations have been studied in literature and applied in order to perform adaptive resampling within sequential Monte Carlo (SMC) methods [11]. For instance, another measure called perplexity, involving the discrete entropy [4] of the normalized weights has been also proposed in [2]; see also [18, Chapter 4], [7, Section 3.5]. Another expression is defined as the inverse of the maximum of the normalized weights $\bar{w}_n$ [17].

In this work, we recall the definition of the Generalized ESS (G-ESS) functions given [17] and show that the G-ESS functions can be considered *diversity indices* [12] (see third box in Figure 1). With this aim, we analyze a G-ESS family (called hereafter *Huggins-Roy's family*) introduced and studied independently in [11] which contains all the main ESS approximations proposed in literature. Moreover, all the functions included in the Huggins-Roy's family [11] fulfill the conditions described in [17] and recalled in this work. We show that this family is related to the Rényi entropy of the probability mass function (pmf) defined by $\bar{w}_n$, $n = 1, \dots, N$ [4]. Furthermore, by numerical simulations, we obtain the G-ESS function within Huggins-Roy's family which provides the best approximation the theoretical ESS definition (in some specific scenarios). We also study linear combinations of G-ESS functions in order to enhance the approximation of the theoretical definition. The results of our numerical simulations suggest the use of the formula $\widehat{ESS} = \left(\frac{1}{\sum_{n=1}^{M} \bar{w}_n^4}\right)^{1/3}$.

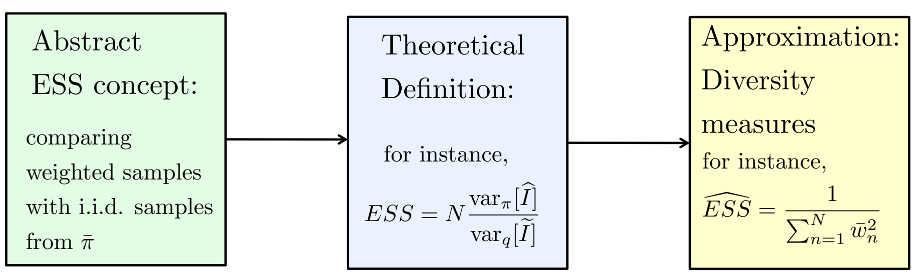


Figure 1: Graphical representation of the development of the approximated ESS formulas for importance sampling. The abstract concept of Effective Sample Size has been translated in a mathematical formulation providing a first attempt of theoretical definition. Since this definition cannot compute, several approximations have been proposed (based only in the information provided by the normalized IS weights). The expression $\widehat{ESS} = \frac{1}{\sum_{n=1}^{M} \bar{w}_n^2}$ is the most applied so far in the literature.

1 Effective Sample Size for Importance Sampling

Let us denote the target probability density function (pdf) as $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$ (known up to a normalizing constant) with $\mathbf{x} \in \mathcal{X}$. Moreover, we consider the following integral involving $\bar{\pi}(\mathbf{x})$

and a square-integrable function $h(\mathbf{x})$,

$$I = \int_{\mathcal{X}} h(\mathbf{x})\bar{\pi}(\mathbf{x})d\mathbf{x}, \tag{1}$$

which we desire to approximate using a Monte Carlo approach. If we are able to draw N independent samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from $\bar{\pi}(\mathbf{x})$, then the Monte Carlo estimator of I is

$$\widehat{I} = \frac{1}{N}\sum_{n=1}^{N} h(\mathbf{x}_n) \approx I, \tag{2}$$

where $\mathbf{x}_n \sim \bar{\pi}(\mathbf{x})$, with $n = 1, \ldots, N$. However, in general, generating samples directly from the target, $\bar{\pi}(\mathbf{x})$, is impossible. Alternatively, we can draw N samples $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from a (simpler) proposal pdf $q(\mathbf{x})$,[1] and then assign a weight to each sample, $w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$, with $n = 1, \ldots, N$, according to the importance sampling (IS) approach. Defining the normalized weights,

$$\bar{w}_n = \frac{w_n}{\sum_{i=1}^{N} w_i}, \quad n = 1, \ldots, N, \tag{3}$$

then the self-normalized IS estimator is

$$\widetilde{I} = \sum_{n=1}^{N} \bar{w}_n h(\mathbf{x}_n) \approx I, \tag{4}$$

with $\mathbf{x}_n \sim q(\mathbf{x})$, $n = 1, \ldots, N$. In general, the estimator $\widetilde{I}$ is less efficient than $\widehat{I}$, since the samples are not directly drawn from $\bar{\pi}(\mathbf{x})$. In several applications [5, 6], it is necessary to measure the loss of the efficiency using $\widetilde{I}$ instead of $\widehat{I}$. The idea is to define the Effective Sample Size (ESS) as the ratio of the variances of the estimators [13],

$$ESS = N\frac{\mathrm{var}_\pi[\widehat{I}]}{\mathrm{var}_q[\widetilde{I}]}. \tag{5}$$

Finding a useful expression of ESS derived analytically from the theoretical definition is not straightforward. Then, different derivations [13, 14], [6, Chapter 11], [18, Chapter 4] proceed using several approximations and assumptions for yielding an expression useful from a practical point of view. A well-known rule of thumb, widely used in literature [6, 16, 18], is

$$\widehat{ESS}_N(\bar{\mathbf{w}}) = \frac{1}{\sum_{i=1}^{N} \bar{w}_n^2} = \frac{\left(\sum_{i=1}^{N} w_n\right)^2}{\sum_{i=1}^{N} w_n^2}, \tag{6}$$

where we have used the the normalized weights

$$\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N],$$

[1]We assume that $q(\mathbf{x}) > 0$ for all $\mathbf{x}$ where $\bar{\pi}(\mathbf{x})) \neq 0$, and $q(\mathbf{x})$ has heavier tails than $\bar{\pi}(\mathbf{x})$.

in the first equality, and the unnormalized ones in the second equality.[2] An interesting property of the expression Eq. (6) is

$$1 \leq \widehat{ESS}_N(\bar{\mathbf{w}}) \leq N. \tag{7}$$

Another similar measure, called *perplexity*, has been proposed in literature [2, 18] based only on the normalized importance weights,

$$\widehat{ESS}_N(\bar{\mathbf{w}}) = \exp\{H(\bar{\mathbf{w}})\} \tag{8}$$

where

$$H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log \bar{w}_n$$

is the discrete entropy of the vector $\bar{\mathbf{w}}$ [4]. Note that again $1 \leq \widehat{ESS}_N(\bar{\mathbf{w}}) \leq N$. In the following, we describe five conditions that a generic ESS approximation based only on the information of the normalized weights must satisfy. Then we show that the family of functions proposed in [11] fulfills these five conditions. Furthermore, we link this G-ESS family with the Rényi entropy providing also some theoretical results.

2 Generalized ESS functions

Here, we recall five properties that a generalized ESS measure (G-ESS) should satisfy, based only on the information of the normalized weights. Here, first of all, note that any possible G-ESS is a function of the vector of normalized weights $\bar{\mathbf{w}} = [\bar{w}_1, \dots, \bar{w}_N]$,

$$E_N(\bar{\mathbf{w}}) = E_N(\bar{w}_1, \dots, \bar{w}_N) : \mathcal{S}_N \to [1, N], \tag{9}$$

where $\mathcal{S}_N \subset \mathbb{R}^N$ represents the *unit simplex* in $\mathbb{R}^N$. Namely, the variables $\bar{w}_1, \dots, \bar{w}_N$ are subjected to the constrain

$$\bar{w}_1 + \bar{w}_2 + \ldots + \bar{w}_N = 1. \tag{10}$$

Moreover, we denoted

$$\bar{\mathbf{w}}^* = \left[\frac{1}{N}, \dots, \frac{1}{N}\right], \tag{11}$$

and the vertices of the simplex $\mathcal{S}_N$ are denoted as

$$\bar{\mathbf{w}}^{(j)} \quad = \quad [\bar{w}_1 = 0, \dots, \bar{w}_j = 1, \dots, \bar{w}_N = 0], \tag{12}$$

i.e., $\bar{w}_j = 1$ and $\bar{w}_n = 0$ (it can occurs only if $\pi(\mathbf{x}_n) = 0$), for $n \neq j$ with $j \in \{1, \dots, N\}$.

Below we list the five conditions that $E_N(\bar{\mathbf{w}})$ should fulfill:

[2] Due to the several approximations which have been applied to obtain the final formula, P_N does not depend on the particles $\mathbf{x}_n$, $n = 1, \dots, N$, which is obviously a drawback (for further considerations see [17]).

C1. **Symmetry:** E_N must be invariant under any permutation of the weights, i.e.,

$$E_N(\bar{w}_1, \bar{w}_2, \dots, \bar{w}_N) = E_N(\bar{w}_{j_1}, \bar{w}_{j_2}, \dots, \bar{w}_{j_N}), \tag{13}$$

for any possible set of indices $\{j_1, \dots, j_N\} = \{1, \dots, N\}$.

C2. **Maximum condition:** A maximum value is N and it is reached at $\bar{\mathbf{w}}^*$ (see Eq. (11)), i.e.,

$$E_N(\bar{\mathbf{w}}^*) = N \geq E_N(\bar{\mathbf{w}}). \tag{14}$$

C3. **Minimum condition**: the minimum value is 1 and it is reached (at least) at the vertices $\bar{\mathbf{w}}^{(j)}$ of the unit simplex in Eq. (12),

$$E_N(\bar{\mathbf{w}}^{(j)}) = 1 \leq E_N(\bar{\mathbf{w}}). \tag{15}$$

for all $j \in \{1, \dots, N\}$.

C4. **Unicity of extreme values:** The maximum at $\bar{\mathbf{w}}^*$ is unique and the the minimum value 1 is reached *only* at the vertices $\bar{\mathbf{w}}^{(j)}$, for all $j \in \{1, \dots, N\}$.

C5. **Scalability:** Consider the vector of weights $\bar{\mathbf{w}} \in \mathbb{R}^N$ and the vector $\bar{\mathbf{v}} = [\bar{v}_1, \dots, \bar{v}_{MN}] \in \mathbb{R}^{MN}$, $\quad M \geq 1$, obtained repeating and scaling by $\frac{1}{M}$ the entries of $\bar{\mathbf{w}}$, i.e.,

$$\bar{\mathbf{v}} = \frac{1}{M}[\underbrace{\bar{\mathbf{w}}, \bar{\mathbf{w}}, \dots, \bar{\mathbf{w}}}_{M-times}]. \tag{16}$$

The invariance condition is expressed as

$$E_N(\bar{\mathbf{w}}) \quad = \quad \frac{1}{M} E_{MN}(\bar{\mathbf{v}}), \tag{17}$$

for all $M \in \mathbb{N}^+$.

This last requirement can be interpreted as an adjustment of the well-known *homogeneity* (scale-invariance) condition for real functions.[3]

3 G-ESS functions as diversity measure

The Huggins-Roy's family introduced in [11] is defined as

$$H_N^{(\beta)}(\bar{\mathbf{w}}) \quad = \quad \left(\frac{1}{\sum_{n=1}^N \bar{w}_n^\beta} \right)^{\frac{1}{\beta-1}}, \tag{18}$$

$$= \quad \left(\sum_{n=1}^N \bar{w}_n^\beta \right)^{\frac{1}{1-\beta}}, \qquad \beta \geq 0. \tag{19}$$

[3]A function $f(\mathbf{x})$ is said to be homogeneous of degree k if $f(c\,\mathbf{x}) = c^k f(\mathbf{x})$ where c is a non-zero constant value.

Table 1 shows that the Huggins-Roy's family contains all the main G-ESS functions introduced in literature. The special cases with $\beta = 0$ and $\beta = 1$ bring to two undetermined expressions that will be solved and clarified below (when the relationship with Rényi entropy is shown). We can easily note that $1 \leq H_N^{(\beta)}(\bar{\mathbf{w}}) \leq N$ for all $\beta \geq 0$. More generally, it is straightforward to observe that the conditions C1, C2, C3 and C4 are fulfilled (with the exception of $\beta = 0$ that does not satisfy C4). In order to prove the condition C5, for simplicity let us consider a vector $\bar{\mathbf{v}} = \frac{1}{2}[\bar{\mathbf{w}}, \bar{\mathbf{w}}]$, defined repeating twice the vector $\bar{\mathbf{w}}$ (i.e., $M = 2$). In this case, we have

$$
\begin{aligned}
H_{2N}^{(\beta)}(\bar{\mathbf{v}}) &= \left(\frac{1}{2^\beta}\sum_{n=1}^{N}\bar{w}_n^\beta + \frac{1}{2^\beta}\sum_{n=1}^{N}\bar{w}_n^\beta\right)^{\frac{1}{1-\beta}}, && (20)\\
&= \left(\frac{1}{2^{\beta-1}}\sum_{n=1}^{N}\bar{w}_n^\beta\right)^{\frac{1}{1-\beta}}, && (21)\\
&= 2\left(\sum_{n=1}^{N}\bar{w}_n^\beta\right)^{\frac{1}{1-\beta}}, && (22)\\
&= 2H_N^{(\beta)}(\bar{\mathbf{w}}), && (23)
\end{aligned}
$$

which is exactly the condition in Eq. (17). Note that other G-ESS families have been proposed [17], but only certain functions (specific cases) fulfill the 5 conditions described above.

Table 1: Special cases of G-ESS functions contained in the Huggins-Roy's family.

$\beta = 0$	$\beta = 1/2$	$\beta = 1$	$\beta = 2$	$\beta = \infty$
$N - N_Z$ *where N_Z is the number of zeros in $\bar{\mathbf{w}}$*	$\left(\sum_{n=1}^{N}\sqrt{\bar{w}_n}\right)^2$	$\exp\left(-\sum_{n=}^{N}\bar{w}_n\log\bar{w}_n\right)$ *(perplexity)*	$\frac{1}{\sum_{n=1}^{N}\bar{w}_n^2}$ *(standard approximation)*	$\frac{1}{\max[\bar{w}_1,\ldots,\bar{w}_N]}$

3.1 Relationship with the Rényi entropy

The Rényi entropy [4] is defined as

$$
R_N^{(\beta)}(\bar{\mathbf{w}}) = \frac{1}{1-\beta}\log\left[\sum_{n=}^{N}\bar{w}_n^\beta\right], \qquad \beta \geq 0, \tag{24}
$$

(25)

Then, it is straightforward to note that

$$
H_N^{(\beta)}(\bar{\mathbf{w}}) = \exp\left(R_N^{(\beta)}(\bar{\mathbf{w}})\right), \tag{26}
$$

i.e., the Huggins-Roy's family contains *diversity indices* derived by the Rényi entropy [4, 12]. For $\beta = 0$, we have $R_N^{(0)}(\bar{\mathbf{w}}) = \log(N - N_Z)$ [4] where $N_Z = \#\{\text{all } \bar{w}_n\text{:} \quad \bar{w}_n = 0, \quad \forall n = 1, \ldots, N\}$, so that $H_N^{(0)}(\bar{\mathbf{w}}) = N - N_Z$. For $\beta = 1$, we have $R_N^{(0)}(\bar{\mathbf{w}}) = -\sum_{n=}^{N} \bar{w}_n \log \bar{w}_n$ [4] then

$$H_N^{(1)}(\bar{\mathbf{w}}) = \exp\left(-\sum_{n=}^{N} \bar{w}_n \log \bar{w}_n\right), \tag{27}$$

The connection with the Rényi entropy shows that the G-ESS functions contained in the Huggins-Roy's family are diversity indices [12]. Moreover, this observation allow us to obtain some theoretical results about $H_N^{(\beta)}$. Indeed, for instance, it is well-known that [4]

$$R_N^{(0)}(\bar{\mathbf{w}}) \geq R_N^{(1)}(\bar{\mathbf{w}}) \geq R_N^{(2)}(\bar{\mathbf{w}}) \geq \ldots R_N^{(\beta')}(\bar{\mathbf{w}}) \ldots \geq R_N^{(\infty)}(\bar{\mathbf{w}}),$$

with $\beta' \geq 2$, Then, since $H_N^{(\beta)}$ is an increasing monotonic function of $R_N^{(\beta)}$, we can also assert

$$H_N^{(0)}(\bar{\mathbf{w}}) \geq H_N^{(1)}(\bar{\mathbf{w}}) \geq H_N^{(2)}(\bar{\mathbf{w}}) \ldots H_N^{(\beta')}(\bar{\mathbf{w}}) \ldots \geq H_N^{(\infty)}(\bar{\mathbf{w}}). \tag{28}$$

with $\beta' \geq 2$. Moreover, since [4]

$$R_N^{(2)}(\bar{\mathbf{w}}) \leq 2R_N^{(\infty)}(\bar{\mathbf{w}}),$$

we also have

$$H_N^{(2)}(\bar{\mathbf{w}}) \leq 2H_N^{(\infty)}(\bar{\mathbf{w}}). \tag{29}$$

4 Numerical Simulations

First of all, we recall the theoretical definition of ESS in Eq. (5),

$$ESS(h) = N\frac{\text{var}_\pi[\widehat{I}]}{\text{var}_q[\widetilde{I}]}. \tag{30}$$

where we consider $x \in \mathbb{R}$ the use of the integrand $h(x) = x$ (in the definition above, we have clarified the dependence on the function h). Namely, $\widehat{I}$ and $\widetilde{I}$ are estimators of the expected value of a random variable X with pdf $\bar{\pi}(x)$. In this numerical example, we compute approximately via Monte Carlo the theoretical definition ESS and we compare them with the G-ESS functions $H_N^{(\beta)}$. More specifically, we consider a univariate standard Gaussian density as target pdf,

$$\bar{\pi}(x) = \mathcal{N}(x; 0, 1), \tag{31}$$

and also a Gaussian proposal pdf,

$$q(x) = \mathcal{N}(x; \mu_p, \sigma_p^2), \tag{32}$$

with mean μ_p and variance σ_p^2. We set $\sigma_p = 1$ and vary $\mu_p \in [0, 2]$. Figure 2 depicts this experimental setup for two specific values of μ_p, 0.5 and 1.5. Clearly, for $\mu_p = 0$ we have the ideal Monte Carlo case, $q(x) \equiv \bar{\pi}(x)$. As μ_p increases, the proposal becomes more different from $\bar{\pi}$. We set $N = 1000$.

Figure 3 shows the theoretical ESS/N curves (solid line) $H_N^{(2)}/N$ (circles) and $H_N^{(\infty)}/N$ (squares), averaged over 10^5 independent runs. Note that $\frac{1}{N} \leq \frac{ESS}{N} \leq 1$.

Optimal linear combination of $H_N^{(2)}$ and $H_N^{(\infty)}$. The functions $H_N^{(2)}$ and $H_N^{(\infty)}$ are the most promising approximations of the theoretical ESS definition (as confirmed by different studies [11, 17]). For this reason, we also consider the linear combination of the G-ESS formulas $H_N^{(2)}$ and $H_N^{(\infty)}$,

$$E_N(\bar{\mathbf{w}}) = a_1 H_N^{(2)}(\bar{\mathbf{w}}) + a_2 H_N^{(\infty)}(\bar{\mathbf{w}}). \tag{33}$$

This example suggests the use of

$$\begin{aligned} a_1 &= 0.6245, \\ a_2 &= 0.4289, \end{aligned} \tag{34}$$

obtained using a Least Squares (LS) regression in order to obtain an expression $E_N(\bar{\mathbf{w}})$ as close as possible to the theoretical ESS curve.

Optimal β for $H_N^{(\beta)}(\bar{\mathbf{w}})$. Furthermore, we have computed the curves (as function β) of $H_N^{(\beta)}(\bar{\mathbf{w}})$ for different values of β, considering a thin grid of β values from 0.2 to 50 with a step of 0.01 (i.e., $\beta \in \mathcal{G}$ denoting $\mathcal{G}$ the thin grid). We consider a L_1 distance between each $H_N^{(\beta)}(\bar{\mathbf{w}})$ curve and the theoretical ESS curve,[4] denoted as $D_1(H_N^{(\beta)}, ESS)$, and compute

$$\beta^* = \arg\min_{\beta \in \mathcal{G}} D_1(H_N^{(\beta)}, ESS). \tag{35}$$

With this procedure, we obtain

$$\beta^* \approx 4.$$

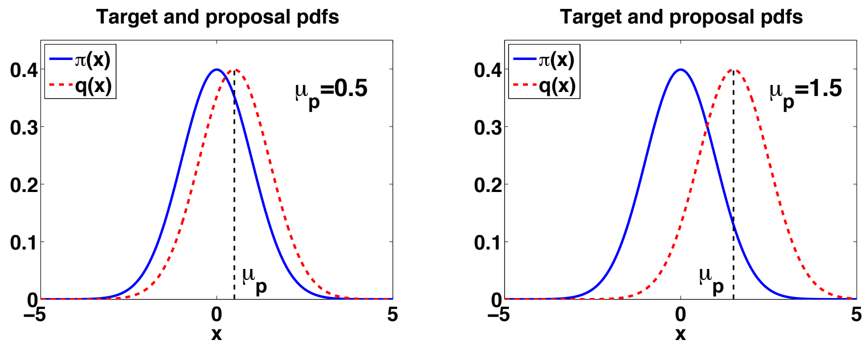
Discussion of the results. Figure 4 shows the curves of the ESS rates corresponding to the theoretical ESS curve (solid line), the best linear combination corresponding to the Eqs. (33)-(34) (squares) and the curve corresponding to $H_N^{(\beta^*)}$ (dashed line). First of all, we can note that the linear combination can return values greater than 1 (recall that we are considering ESS/N) that is clearly a drawback. Moreover, we can see that the curve corresponding to $H_N^{(4)}(\bar{\mathbf{w}})$ fits particularly well in this numerical setup, providing a close very close to the theoretical ESS curve. Observe that the approximation provided by $H_N^{(4)}$ is virtually perfect for $\mu_p \leq 1$. Hence, we suggest the use of the expression

$$H_N^{(4)}(\bar{\mathbf{w}}) \;=\; \left(\frac{1}{\sum_{n=1}^N \bar{w}_n^4}\right)^{\frac{1}{3}}. \tag{36}$$

5 Conclusions

In this work, we have analyzed alternative Effective Sample Size (ESS) measures for Monte Carlo algorithms based on the importance sampling techniques. More specifically, we have studied the

[4] Recall that these curves are functions of μ_p and are averaged over 10^5 independent runs.



(a) (b)

Figure 2: Target and proposal pdfs with $\mu_p = \{0.5, 1.5\}$. The variances of both densities is set to 1.

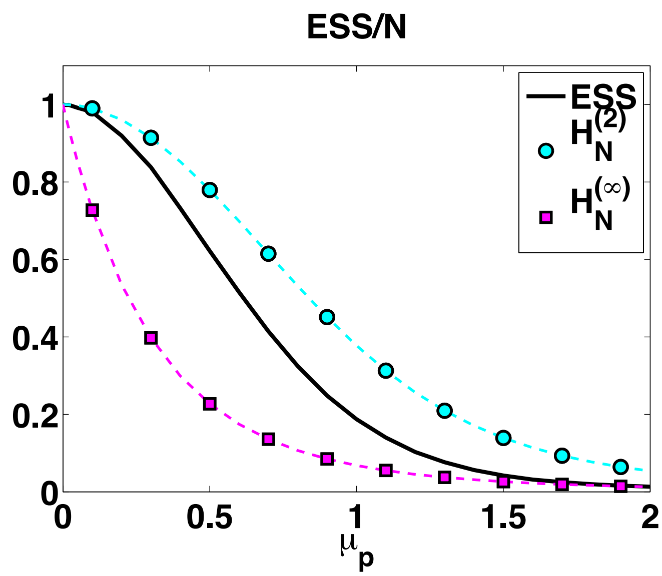


Figure 3: ESS rates (i.e., the ratio of ESS values over N) corresponding to the theoretical ESS value (solid line), $H_N^{(2)}$ (circles) and $H_N^{(\infty)}$ (squares). We set $N = 1000$.

family of ESS approximations introduced in [11]. We have shown that all the ESS functions included in this family (called Huggins-Roy's family) fulfill all the required theoretical conditions described in [17] and we have remarked that they can be interpreted as *diversity indices* [12]. We have also highlighted the relationship of this family with the Rényi entropy [4]. Furthermore, we have studied the performance of different ESS approximations introducing also an optimal linear combination of the most promising ESS indices by numerical simulations. In the numerical example, we have obtained the best ESS approximation within the Huggins-Roy's family, that is $H_N^{(4)}(\bar{\mathbf{w}}) = \left(\frac{1}{\sum_{n=1}^N \bar{w}_n^4}\right)^{\frac{1}{3}}$. This formula provides almost a perfect match with the theoretical ESS values, in the considered experimental setup.

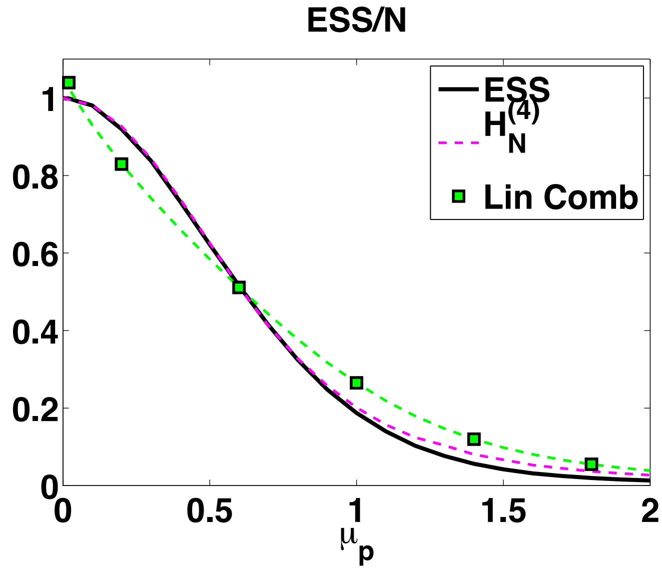


Figure 4: ESS rates (i.e., the ratio of ESS values over N) corresponding to the theoretical ESS value (solid line), $H_N^{(4)}$ (dashed line) and the linear combination E_N in Eq. (33)-(34) (squares). We set $N = 1000$. The approximation provided by $H_N^{(4)}$ is virtually perfect for $\mu_p \leq 1$.

Acknowledgement

Luca Martino acknowledges support by the Agencia Estatal de Investigación AEI (project SPGRAPH, ref. num. PID2019- 105032GB-I00).

References

[1] M. F. Bugallo, L. Martino, and J. Corander. Adaptive importance sampling in signal processing. *Digital Signal Processing*, 47:36–49, 2015.

[2] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert. Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18:447–459, 2008.

[3] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York (USA), 1991.

[5] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, September 2003.

[6] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice.* Springer, New York, 2001.

[7] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *technical report*, 2008.

[8] V. Elvira, L. Martino, and C. P. Robert. Rethinking the effective sample size. *arXiv:1809.04129*, 2018.

[9] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference,*. Chapman & Hall/CRC Texts in Statistical Science, 2006.

[10] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proceedings-F Radar and Signal Processing*, 140:107–113, 1993.

[11] J. H Huggins and D. M Roy. Convergence of sequential Monte Carlo based sampling methods. *arXiv:1503.00966*, 2015.

[12] L. Jost. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

[13] A. Kong. A note on importance sampling using standardized weights. *Technical Report 348, Department of Statistics, University of Chicago*, 1992.

[14] A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

[15] F. Liang, C. Liu, and R. Caroll. *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples.* Wiley Series in Computational Statistics, England, 2010.

[16] J. S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, 2004.

[17] L. Martino, V. Elvira, and F. Louzada. Effective sample size for importance sampling based on discrepancy measures. *Signal Processing*, 131:386–401, 2017.

[18] C. P. Robert and G. Casella. *Introducing Monte Carlo Methods with R.* Springer, 2010.